Year: 2020

Effects of fundamental frequency, vocal intensity, sample duration, and vowel context in cepstral and spectral measures of dysphonic voices

Carvalho Sampaio, Marilia ; Vaz Masson, M ; de Paula, Soares ; Bohlender, J E ; Brockmann-Bauser, Meike

Abstract: Purpose Smoothed cepstral peak prominence (CPPS) and harmonics-to-noise ratio (HNR) are acoustic measures related to the periodicity, harmonicity, and noise components of an acoustic signal. To date, there is little evidence about the advantages of CPPS over HNR in voice diagnostics. Recent studies indicate that voice fundamental frequency (F0) and intensity (sound pressure level [SPL]), sample duration (DUR), vowel context (speech vs. sustained phonation), and syllable stress (SS) may influence CPPS and HNR results. The scope of this work was to investigate the effects of voice F0 and SPL, DUR, SS, and token on CPPS and HNR in dysphonic voices. Method In this retrospective study, 27 Brazilian Portuguese speakers with voice disorders were investigated. Recordings of sustained vowels (SVs) /a:/ and manually extracted vowels (EVs) /a/ from Consensus Auditory-Perceptual Evaluation of Voice sentences were acoustically analyzed with the Praat program. Results There was a highly significant effect of F0, SPL, and DUR on both CPPS and HNR ($p < .001$), whereas SS and vowel context significantly affected CPPS only ($p < .05$). Higher SPL, F0, and lower DUR were related to higher CPPS and HNR. SVs moderately-to-highly correlated with EVs for CPPS, whereas HNR had few and moderate correlations. In addition, CPPS and HNR highly correlated in SVs and seven EVs ($p < .05$). Conclusion Speaking prosodic variations of F0, SPL, and DUR influenced both CPPS and HNR measures and led to acoustic differences between sustained and excised vowels, especially in CPPS. Vowel context, prosodic factors, and token type should be controlled for in clinical acoustic voice assessment.

DOI: https://doi.org/10.1044/2020_JSLHR-19-00049

Originally published at:
Carvalho Sampaio, Marilia; Vaz Masson, M; de Paula, Soares; Bohlender, J E; Brockmann-Bauser, Meike (2020). Effects of fundamental frequency, vocal intensity, sample duration, and vowel context in cepstral and spectral measures of dysphonic voices. Journal of Speech, Language, and Hearing Research, 63(5):1326-1339.
DOI: https://doi.org/10.1044/2020_JSLHR-19-00049

Effects of Fundamental Frequency, Vocal Intensity, Sample Duration, and Vowel Context in Cepstral and Spectral Measures of Dysphonic Voices

Marília Sampaio[1,2], Maria Lúcia Vaz Masson[1], Maria Francisca de Paula Soares[1], Jörg Edgar Bohlender [2,3], Meike Brockmann-Bauser[2,3]

[1] Department of Speech-Language and Hearing Sciences, Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil.

[2] Department of Phoniatrics and Speech Pathology, Clinic for Otorhinolaryngology Head and Neck Surgery, University Hospital Zurich, Zurich, Switzerland.

[3]University of Zurich, Zurich, Switzerland.

Corresponding Author: Marília Sampaio

Department of Speech-Language and Hearing Sciences, Institute of Health Sciences, Federal University of Bahia. Av Reitor Miguel Calmon, s/n, 1º andar, Vale do Canela, CEP 40.110-100 Salvador, Bahia, Brazil. +55 71 3283-8886, +55 71 98858-2017, marilia.sampaio@ufba.br

Abstract

Purpose: Smoothed cepstral peak prominence (CPPS) and harmonics-to-noise ratio (HNR) are acoustic measures related to the periodicity, harmonicity, and noise components of an acoustic signal. To date, there is little evidence about the advantages of CPPS over HNR in voice diagnostics. Recent studies indicate that voice fundamental frequency (F0) and intensity (SPL), sample duration (DUR), vowel context (speech versus sustained phonation), and syllable stress (SS) may influence CPPS and HNR results. The scope of this work was to investigate the effects of voice F0 and SPL, DUR, SS, and token on CPPS and HNR in dysphonic voices. Methods: In this retrospective study, 27 Brazilian Portuguese speakers with voice disorders were investigated. Recordings of sustained vowels (SV) /a:/ and manually extracted vowels (EV) /a/ from CAPE-V sentences were acoustically analyzed with PRAAT program. Results: There was a highly significant effect of F0, SPL, and DUR on both CPPS and HNR ($p<.001$), while SS and vowel context significantly affected CPPS only ($p<.05$). Higher SPL, F0, and lower DUR were related to higher CPPS and HNR. SV moderate-to-highly correlated with EVs for CPPS, while HNR had few and moderate correlations. In addition, CPPS and HNR highly correlated in SV and seven EVs ($p<.05$). Conclusion: Speaking prosodic variations of F0, SPL, and DUR influenced both CPPS and HNR measures and led to acoustic differences between sustained and excised vowels, especially in CPPS. Vowel context, prosodic factors, and token type should be controlled for in clinical acoustic voice assessment.

Keywords: voice disorders, speech acoustics, acoustic analysis, dysphonia, CPPS

Effects of Fundamental Frequency, Vocal Intensity, Sample Duration, and Vowel Context in Cepstral and Spectral Measures of Dysphonic Voices

Introduction

Acoustic voice analysis is a widely applied quantitative and noninvasive instrument for the clinical evaluation of vocal function. The development of affordable recording methodologies and robust acoustic measures has encouraged its use in varied clinical and research applications (Maryn & Weenink, 2015; Patel et al., 2018; Vogel & Maruff, 2008; Watts, Awan, & Maryn, 2017). Recently, cepstral peak prominence (CPP) was recommended in a standardized voice assessment protocol as a general measure of dysphonia (Patel et al., 2018). However, despite being proposed as the sole measure for vocal quality, there is still a lack of evidence about the advantages in the clinical application of CPP over other measures, such as harmonics-to-noise ratio (HNR) (Aichinger & Kubin, 2018; Awan, Awan, Watts, & Gaskill, 2018; Riesgo & Nöth, 2019; Vaz Freitas, Melo Pestana, Almeida, & Ferreira, 2015).

What do cepstral peak prominence and harmonics-to-noise ratio indicate?

A cepstrum has been described as “a log power spectrum of a log power spectrum” (Hillenbrand & Houde, 1996). CPP measures the amplitude of the cepstral peak (the first rahmonic) compared to the resembling value on the regression line that runs through the cepstrum directly below the first peak. The time of the first cepstral peak corresponds to the fundamental period. Besides, the prominence of the cepstral peak is a measure of how sinusoidal the power spectrum of the signal is (Riesgo & Nöth, 2019). Thus, a signal whose spectrum shows a clear harmonic structure will show a very prominent cepstral peak. An alternative cepstral measure is smoothed cepstral peak prominence (CPPS), with an additional processing step of smoothing the cepstra before calculating the peak prominence (Hillenbrand & Houde, 1996).

As a comparison, the spectrum-based HNR evaluates the degree of the acoustic periodicity of the voice signal by quantifying the ratio between the periodic (harmonic part) and aperiodic (noise) components. HNR estimation may be based on the frequency representation through Fourier transformation or may depend on a time-based parameter, such as in the cross-correlation method in Praat program (Boersma, 1993; Teixeira & Fernandes, 2015).

CPPS and HNR are both generally used to quantify perceptual dysphonia in several laryngeal conditions (Madill, Nguyen, Cham, Novakovic, & McCabe, 2019). Since CPP, CPPS, and HNR are related to periodicity, harmonicity, and noise components, they may correlate under certain conditions. This has been reported in breathy phonations, with rising of the noise floor in the speech spectrum (Heman-Ackah, Michael, & Goding, 2002; Shue, Chen, & Alwan, 2010). CPP and CPPS have been described as more reliable in dysphonic voices, since they do not rely on correct recognition of fundamental frequency (F0) and sound pressure level (SPL) to the same extent as so-called traditional acoustic measures such as jitter and shimmer (Halberstam, 2004; Heman-Ackah et al., 2002; Patel et al., 2018). Moreover, CPP and CPPS showed high correlation and accuracy indices for detection of perceptual dysphonia in sustained vowels and also in connected speech (Awan, Roy, Jette, Meltzner, & Hillman, 2010; Brinca, Batista, Tavares, Goncalves, & Moreno, 2014; Delgado, León, Jiménez, & Izquierdo, 2017; Delgado-Hernandez, Leon-Gomez, Izquierdo-Arteaga, & Llanos-Fumero, 2018; Halberstam, 2004; Hasanvand, Salehi, & Ebrahimipour, 2017; Heman-Ackah et al., 2003; Moers et al., 2012; Nuñez-Batalla et al., 2018; Phadke et al., 2018; Selamtzis, Castellana, Salvi, Carullo, & Astolfi, 2019). Even though CPP has been described as superior to HNR (Riesgo & Nöth, 2019), some works also show a strong and significant correspondence between HNR and perceptual analysis of sustained vowels as a result of improvements in HNR estimation methods (Moers et al., 2012; Vaz Freitas et al., 2015). Further, a recent study showed that CPP is sensitive to changes in vocal tract configuration, while HNR is less affected by resonatory conditions, and may be more reliable than CPP in documenting dysphonia related to signal aperiodicity in sustained vowels (Madill et al., 2019). Thus, the benefits of CPP over HNR in the daily practice of voice assessment are still open to discussion (Aichinger & Kubin, 2018; Awan et al., 2018).

Considering that CPPS is less dependent on correct F0 and SPL estimation in short term analysis, we expect in our work that CPPS presents a better performance compared to HNR. In turn, we also expect that HNR will be less influenced by prosodic effects compared to CPPS.

Sustained vowels versus speech samples in cepstral and spectral-based analysis

Currently, the debate about the ecological validity of sustained vowel (SV) versus speech samples in clinical voice evaluation has led to controversial results about the most appropriate sample for acoustic analysis (Gerratt, Kreiman, & Garellek, 2016). For many years, vowel phonations have been favored over speech tasks in clinical measurements, due to its steady properties and higher relative reliability in the calculation of perturbation or spectral-based indices (Halberstam, 2004; Heman-Ackah et al., 2003; Heman-Ackah et al., 2002; Moers et al., 2012; Parsa & Jamieson, 2001; Zhang & Jiang, 2008). In more recent studies, the advantage of speech samples for its more natural communication context has been described, representing a great benefit to voice diagnostics (Delgado-Hernandez et al., 2018; Heman-Ackah et al., 2003; Heman-Ackah et al., 2014; Kitayama et al., 2018; Maryn & Weenink, 2015; Phadke et al., 2018; Sauder, Bretl, & Eadie, 2017; Selamtzis et al., 2019; Watts et al., 2017).

However, for CPP and CPPS the results regarding sustained vowels and connected speech were contradictory. While four studies found significant differences between healthy and dysphonic voices for both SV and speech samples (Brinca et al., 2014; Delgado-Hernandez et al., 2018; Hasanvand et al., 2017; Selamtzis et al., 2019), others reported higher CPP or CPPS values for SV as compared to speech (Kitayama et al., 2018; Phadke et al., 2018). Further, Gerrat et al. (2016) reported no significant differences between SV and speech. Also, correlation tests of CPPS showed a high agreement between SV and speech scores in dysphonic voices, and weak to moderate correlation in healthy voices (Nuñez-Batalla et al., 2018).

Furthermore, the agreement of token type with perceptual dysphonia has been described in contradictory terms. In several studies, CPP or CPPS from speech samples correlated better to perceptual dysphonia as compared to SV samples (Delgado-Hernandez et al., 2018; Halberstam, 2004; Moers et al., 2012), whereas in other works SV was better or equally correlated with dysphonia (Awan et al., 2010; Brinca et al., 2014; Heman-Ackah et al., 2002), depending on the perceptual parameter. In tests of accuracy, both CPP and CPPS derived from SV and vowels from connected speech showed a similar discriminatory power for the detection of perceptual dysphonia (Heman-Ackah et al., 2003; Watts & Awan, 2011). However, Selamtzis et al. (2019) found a better discriminatory power in extracted vowels from speech.

In the cepstral analysis of connected speech samples, usually complete sentences with heterogeneous information are assessed. These samples may include pauses and unvoiced signals (which may act as noise) or preserve only voiced signals, keeping the fluctuations in intonation and intensity (Brinca et al., 2014; Delgado-Hernandez et al., 2018; Hasanvand et al., 2017; Heman-Ackah et al., 2014; Kitayama et al., 2018; Lowell & Hylkema, 2016; Maryn & Weenink, 2015; Nuñez-Batalla et al., 2018; Phadke et al., 2018; Sauder et al., 2017; Watts et al., 2017). Despite the contributions of this approach to dysphonia detection, the use of entire speech, blended or not with SV, may dissipate the acoustic information and thereby foster inaccurate conclusions about vocal function (Kitayama et al., 2018). In addition, studies using extracted vowels from connected speech not often control for vowel type and utterance position, which might affect measurements results (Castellana, Carullo, Astolfi, Bisetti, & Colombini, 2018; Gerratt et al., 2016; Selamtzis et al., 2019; Zhang & Jiang, 2008).

As for HNR, contradictory results regarding the validity of token type have been reported, and several alternative algorithms such as signal-to-noise ratio (SNR) (Klingholtz, 1990; Parsa & Jamieson, 2001; Zhang & Jiang, 2008), noise-to-harmonics ratio (NHR) (Gerratt et al., 2016; Halberstam, 2004; Heman-Ackah et al., 2003; Heman-Ackah et al., 2002), and harmonics-to-noise ratio (HNR) have been applied (Moers et al., 2012; Parsa & Jamieson, 2001). Zhang & Jiang (2008) found significant differences between SNR measures of healthy versus dysphonic voices for both SV and speech samples. However, SNR measures from speech correlated better with dysphonia severity as compared to SV in Klingholtz's (1990) study. In turn, Gerratt et al., (2016) did not find significant differences between SV and extracted vowels from speech using NHR. In contrast, NHR measures from SV correlated better than speech with auditory-perceptual dysphonia in a study by Halberstam (2004).

The reported inconsistencies in previous studies may be influenced by several factors, such as different recording conditions, software programs and algorithms, severity and type of perceptual dysphonia, F0, SPL, sample duration, and vowel context. To date, the role of these multiple influencing factors remains unclear in the clinical application of CPPS and HNR.

Influence of speech characteristics on cepstral and spectral parameters

Differences in acoustic information of SV versus speech samples have been mainly attributed to prosodic variations of speech, which are represented by aspects such as pitch and intensity contour, stressing patterns, and syllable duration. Prosodic variations are associated with language pragmatics and dialect, personal, educational, cultural, and social aspects (Shevchenko, 2003). For instance, according to Shevchenko (2003), Russian women speak faster, louder, with higher pitch but less resonance as compared to British women, while American women speak with higher pitch and lower accentual duration compared to the British. Further, sociocultural trends are linked to intensity, resonance, and time duration patterns, and contrast between accented and non-accented syllables (Shevchenko, 2003).

Changes in acoustic features are expected to naturally and frequently happen in speech samples of healthy voices due to shifting in voice F0 and SPL, syllable stress, transitions from glides or consonants, and rapid voice onsets/offsets (Awan, Giovinco, & Owens, 2012; Awan et al., 2010; Maryn & Weenink, 2015; Parsa & Jamieson, 2001). Variations in voice F0 and SPL have been mainly related to adjustments in articulation, vocal fold closure and tonus (Awan et al., 2012; Brockmann, Drinnan, Storck, & Carding, 2011; Brockmann-Bauser, Bohlender, & Mehta, 2018; Kumar, Bhat, & Prasad, 2010; Teixeira & Fernandes, 2015; Wolfe & Martin, 1997a, 1997b). Therefore, these variations may affect acoustic measures of signal regularity, stability, and harmonic organization. Further, the interaction between F0, window type, and length might influence the calculation of CPPS (Fraile & Godino-Llorente, 2014; Riesgo & Nöth, 2019; Skowronski, Shrivastav, & Hunter, 2015) and HNR (Boersma, 1993), and consequently assessment results.

The type of syllable stress leads to modifications in F0, SPL, and sample duration, and reflects patterns of vocal behavior, carrying information about vocal tract configuration, laryngeal and respiratory functions (Awan et al., 2010; Gerratt et al., 2016; Parsa & Jamieson, 2001). Despite representing an informative characteristic to explore, syllable stress is rarely investigated in studies about acoustic voice measures of speech utterances. Therefore, the systematic assessment of F0, SPL, sample duration, and syllable stress may enhance our understanding of how these features affect clinical acoustic measurements.

Main research aims

The main aims of our work in dysphonic Brazilian Portuguese native speakers were to a) compare CPPS and HNR measurements between SV and extracted vowels from connected speech (EV); b) investigate the effects of F0, SPL, sample duration and syllable stress on CPPS and HNR in both vowel contexts.

Methods

Study design and subjects

In this cross-sectional retrospective study, 27 adult Brazilian Portuguese native speakers (16 women and 11 men) with a mean age of 46 years (*SD* = 14) were investigated. The mean age was 49 years (*SD* = 13, Range 32-77) for women and 41 years (*SD* = 13, Range 30-70) for men.

Institutional clinic reports and voice recordings of patients with voice disorders or vocal complaints were investigated. Patients were assessed at the Speech-Language Pathology Section of a Brazilian University Hospital between February of 2015 and June of 2018. All participants lived in the state of Bahia and were native speakers of Brazilian Portuguese with a similar dialect. Voice assessments consisted of the history of vocal complaints, laryngoscopy, self-assessment questionnaires, and auditory-perceptual analysis of vocal quality. Laryngeal diagnoses included peripheral or central neurogenic (22%; vocal fold paralysis, spasmodic dysphonia, Parkinson’s disease), structural (70%; nodular lesions, polyps, papilloma, sulcus, chronic inflammation, edema, and post-phono surgery) and functional voice disorders (8%;

vestibular fold hyper constriction, vocal fold-chink) (Rosen & Murry, 2000). When classifying the participant`s professions according to voice use level (do Amaral Catani et al., 2016; Koufman & Isaacson, 1991), 16 participants (59%) were non-vocal, and non-professionals, six (22%) non-vocal and professionals, four (15%) professional voice users, and one (4%) was elite professional voice user. The present work was approved by the Ethics Committee of the Federal University of Bahia (letters n. 2.641.558/ICS and 2.761.949/HUPES).

Perceptual analysis

In order to characterize the participants' vocal quality and degree of voice deviance, a perceptual analysis was performed in the 27 patients' recordings using consecutive sentences of the Brazilian Portuguese Consensus Auditory–Perceptual Evaluation of Voice (CAPE-V) (Behlau, 2004) (Table 1). Three newly trained and one experienced professional performed a preliminary perceptual training and calibration using sentences of ten anonymized patients not included in the study. Then, the anonymized sentences of each patient were played using the same audio device for all raters simultaneously. After all, examiners had performed perceptual evaluations individually, they discussed the disagreements, deciding on a final rating by consensus. The GRB grades from the GBRAS scale (Hirano, 1981) were used, where 0 was considered the absence of deviation, 1 considered as mild, 2 as moderate, and 3 as severe deviation of the specific characteristic.

Perceptual ratings of overall vocal quality by GRBAS scale included mild overall deviance in 13 patients (48%), no overall deviance in three patients (11%), moderate overall deviance in eight patients (30%), and severe overall deviance in three patients (11%). Mild perceptual roughness was found in 13 patients (48%), moderate roughness in nine (33%), and severe roughness in two (8%) participants. Mild breathiness was present in 16 (59%) and moderate to severe breathiness in only two patients (8%). Mild strain was present in three (11%) and moderate in only one (4%).

Voice recording tasks

Voice samples consisted of the sustained vowel (SV) /a:/ and ten extracted /a/ vowels (EVs) from four sentences of the Brazilian Portuguese version of Consensus Auditory–Perceptual Evaluation of Voice (CAPE-V) (Behlau, 2004) produced in the same voice recording session. In the original voice task protocol, participants were asked to say the SVs /a:/, /e:/ and /i:/, the months of the year, the numbers 1 to 20, and the six sentences of the CAPE-V using habitual pitch and SPL. All voice tasks were performed and recorded in this order.

Voice recording techniques

Subjects performed the voice tasks in a silent acoustic booth, with the ambient noise of 31 dB(A), which was 16 dB(A) lower than the quietest human voice intensity registered by the calibrated sound pressure level meter (MLM02 Tacklife© Decibel Meter) in the same place. A monoauricular headset with unidirectional condenser type digital microphone (Satellite AE-216, frequency response range 20-15 KHz, sensitivity 64+/-3dB), positioned at 5cm distance and 45° angle from the mouth was used (Patel et al., 2018; Titze, 1995). Voice recordings were done with Praat program (Boersma & Weenink, 2018), with a 44KHz sampling rate and 32-bit quantization. Samples were saved in .wav-format on a desktop computer. Conversion from uncalibrated voice signal amplitude [as measured by Praat] to calibrated SPL values indicated in dB(A) was achieved using a comparison method (Winholtz & Titze, 1997). For this, calibrated speech weighted noise (Wagener, Kühnel, & Kollmeier, 1999) was recorded with 5 cm distance and 45° angle to the sound source at 50 dB(A), 65 dB(A), 85 dB(A), and 90 dB(A).

The difference between known SPL from the calibrated signal and the measured uncalibrated amplitude values was then calculated and later used to compute calibrated SPL of the voice recordings.

Exclusion criteria

General exclusion criteria were incomplete diagnostic or voice recording files and recordings with incorrectly performed voice tasks. Specific exclusion criteria regarding the acoustic samples were signal-to-noise ratio (SNR, defined as the difference between recording SPL in dBA and ambient SPL in dBA) below 30 dB (Deliyski, Shaw, & Evans, 2005), SV shorter than 1.5 seconds, speech sample duration below 64.7 milliseconds (ms) after excluding transitions, vowel samples with less than five periods (Boersma, 1993), and peak clipping. In order to facilitate the investigation of SPL and F0 effects on cepstral and spectral measures, also type three signals (Titze, 1995) with incorrect and unstable F0 and intensity recognition in Praat were excluded (Boersma & Weenink, 2018).

Voice sample selection and anonymizing

Each acoustic signal was visually inspected for quality, peak clipping, sample rate, token order, and completeness using the software Audacity® version 2.2.2 (Audacity, 2018). With the oscillogram display of Praat, we identified unvoiced sections in the voice recordings and registered the intensity of background noise (dB as indicated by Praat). Further, also in Praat the narrowband spectrogram with the "show pulses" and "show intensity" settings turned on was used for the examination of signal stability. The entire SV and each complete sentence from CAPE-V were saved as separate files with an anonymous code.

Segmentation and manual extraction of the vowel /a/ in connected speech

The online tool WebMaus Basic^{SM} (Kisler, Reichel, & Schiel, 2017; Schiel, 1999) was applied to generate the automatic phonetic transcription and boundaries of each word sentence based on previously provided orthographic information. The Spanish language was chosen considering the similarities with the Portuguese language. Next, all the phonetic transcriptions were reviewed, and the phonetic boundaries of the targeted vowels corrected. For this, we used the wideband spectrogram displayed with the "show pulses" and "show formants," spectrogram's dynamic range set in 35 dB, wave periodicity and amplitude in the oscillogram (Barbosa & Madureira, 2015) and auditory-perceptual word/vowel recognition. Eleven /a/ vowels were extracted from the sentences: Five stressed, five pre-stressed, and one post-stressed (Table 1). Other vowels /a/ were excluded from the analysis due to pragmatic features of the dialect (omission, short time, instability, or low intensity), resulting in the exclusion of the entire first and fourth CAPE-V sentences. Each targeted vowel was manually extracted without the transition part and was saved as a separated file (Castellana, Selamtzis, Salvi, Carullo, & Astolfi, 2017).

Acoustic analysis and outcome measures

Praat program version 6.046 (Boersma & Weenink, 2018) was used to conduct acoustic analysis of the 27 SV and the 202 EV samples. Assessed was the section 0.6 to 1.1 seconds from each recording, applying the same analysis script for both vowel types (Brockmann-Bauser et al., 2018). The mean SNR (Deliyski et al., 2005) of the 229 samples was 43 dB SPL (*SD* = 3).

The selected acoustic measures were time duration of the extracted vowel samples (DUR), mean fundamental frequency (F0), mean sound pressure level (SPL), mean harmonics-to-noise ratio (HNR), and smoothed cepstral peak prominence (CPPS). Table 2 summarizes the technical information and the steps performed in Praat program.

Statistical analysis

Data were analyzed with the software SPSS© version 25 (IBM, 2017). The descriptive analysis comprised mean (M), standard deviation (SD), confidence interval (CI), and frequency distribution through contingency tables and scatter plot graphics.

To test the data distribution, the Shapiro-Wilk test was applied to all 229 vowels. Results showed a non-normal distribution for CPPS, $W = .99$, $p = .03$, and for HNR, $W = .98$, $p = .01$. However, the data distribution between groups, analyzed with the Levene's test of equality of variances, showed homogeneous distribution of CPPS and HNR for syllable stress (respectively $F = 2.44$, $p = .07$ and $F = 1.66$, $p = .18$) and token (respectively $F = 1.73$, $p = .07$ and $F = 1.26$, $p = .25$), which fulfilled the requirements for using parametric tests.

To investigate the suitability of the data for performing regression analysis, assumptions of normality, homoscedasticity, and absence of multicollinearity were tested in residual data. First, the distribution of the residual of HNR and CPPS scores conformed to the diagonal line indicated in the Q-Q plot, showing a normal distribution. Next, the predicted and standardized HNR and CPPS residuals distribution showed random and equal distribution on a scatter plot, confirming homoscedasticity of residuals distribution. Finally, variance inflation factor values showed no correlation between co-variables, $VIF = 1.0 – 2.4$, except for DUR, which showed a weak correlation between SPL, $VIF = 3.7$, and F0, $VIF = 3.6$. Thus, the absence of multicollinearity was confirmed, and the requirements for implementation of regression tests were fulfilled.

To investigate the effects of F0, SPL, DUR, syllable stress and vowel context (SV versus type of EV) on CPPS and HNR, Linear Mixed Model (LMM) analysis with ANCOVA was applied (compound symmetry co-variance type, token vowels as repeated measures, with confidence interval of 95% and significance level ≤ .05. LMM analysis also provided the results of paired mean comparisons between sustained vowels and extracted vowels for CPPS and HNR. After that, estimation of mean values for CPPS and HNR after exclusion of SPL, F0 and DUR effects were calculated; then post hoc paired comparisons between sustained and each extracted vowel was applied (based on estimated means), using the Bonferroni correction to avoid type one errors of multiple comparisons (confidence interval of 95%, significance level ≤ .05).

Moreover, correlation tests between SV and the eleven EVs for CPPS, NHR, F0, and SPL were performed using the Paired sample Pearson coefficient. Additional correlation tests between CPPS and HNR for SV and the eleven EVs were applied using the Spearman`s Rank Correlation Coefficient. Coefficients between 0.1 and 0.3 were considered weak, between 0.4-0.6 moderate, and between 0.7-0.9 strong correlations (Dancey & Reidy, 2007).

Results

Mean differences and correlations between sustained and extracted vowels in CPPS and HNR

Table 3 displays the mean distribution of CPPS, HNR, F0, SPL, and DUR in sustained vowels (SV) and the 11 extracted vowels (EVs). For CPPS, all EVs were significantly different from the sustained vowel (p = .005-.036). In turn, there were no significant differences between sustained and extracted vowels for HNR (p > .05).

Significant correlations between sustained and extracted vowels were more frequent and stronger for CPPS compared to HNR (Table 4). For CPPS, EVs 6 and 8 strongly correlated with SV, r (23-14) =.71-.81, p = .000-.004, and EVs 1, 5, 7, 10 and 11 had moderate to weak correlations, r (15-24) = .45-.59, p = .004-.05. As for HNR, only EVs 2, 5, 6 and 10 showed significant and moderate to weak correlation with SV, r (15-24) = .42-.55, p = .024-.05 (Table 4).

Effects of F0, SPL, DUR, syllable stress, and vowel context on CPPS and HNR

The parameters F0, SPL, and DUR had significant effects on both CPPS ($p \leq .003$) and HNR ($p \leq .02$). All observed effects were more significant for CPPS than for HNR (Table 5). Syllable stress and type of extracted vowel had a significant influence on CPPS ($p \leq .003$), whereas no effect was found for HNR ($p \geq .06$) (Table 5).

Mean fundamental frequency of EVs 1 and 5 significantly and strongly correlated with SV, r (23-24) = .71, p < .001, while EVs 2, 6 and 11 showed a moderate to weak correlation, r (16-23) = .45-.64, p = .003-.05. These five EVs also showed significant correlations for CPPS and HNR (Table 4). The distribution of CPPS and HNR with F0 (Figure 1a) and SPL (Figure 1b) demonstrates an increase of CPPS and HNR with higher F0 and SPL, but the strength of correlation would depend on the vowel context. The distribution of CPPS and HNR with DUR (Figure 2) shows the concentration of higher CPPS and HNR values in lower sample lengths. Additionally, CPPS values had a more gathered distribution, while HNR showed a larger spread configuration (Figure 2). Finally, Figure 3 shows the distribution of CPPS and HNR according to the stressing pattern. The lower acoustic measures are concentrated in the stressed vowels, which also had higher time durations.

Correlation between HNR and CPPS

Spearman´s Rank test showed a strong significant correlation between CPPS and HNR for SV, r (27) = .86, p <.001. As for the extracted vowels, the results showed more variation: EVs 4, 5, 10 and 11 showed a significantly strong correlation between CPPS and HNR, r (14-24) = 0.82-0.86, p <.001, EVs 1, 7 and 9 also showed strong correlations, but with lower values as compared to SV, r (15-22) = 0.70-0.75, p <.001, and EV 2 showed a moderate correlation, r (16) = 0.57, p = .022. There was no significant correlation for CPPS and HNR in EVs 3, 6, and 8.

Estimated means after correction for F0, SPL, and DUR

After adjusting the means and confidence intervals by excluding the confounding factors F0, SPL, and DUR using Bonferroni test, it is possible to observe an inversion in the relationship between sustained and extracted vowels for both CPPS and HNR (Table 6). In CPPS, extracted vowels consistently showed lower means, *Ms* = 22.5-24.4, than SV, M = 30 dB, with only one significant difference between SV and EV5, M = 22.9 p =.04. On the other hand, for HNR, EVs consistently showed higher means, *Ms* = 14.1-19.3, than SV, M = 13 dB, and no significant differences between vowel contexts. The range of CPPS values between EVs narrowed from 4 dB in Table 3 to 2 dB in Table 6. It also happened with HNR values, but the variance between extracted vowels kept broader than in CPPS: from 7 dB in Table 3 to 5 dB in Table 6.

Discussion

In our study of voice disordered adults, speaking prosody variations of F0, SPL, sample duration (DUR) significantly influenced CPPS and HNR measures. These effects may explain the differences between sustained and excised vowels. Further, the vowel context and syllable stress affected CPPS, but not HNR in dysphonic voices. Therefore, vowel context, prosodic factors such as F0 and SPL, and token type should be controlled for in clinical acoustic voice assessment.

Natural prosodic speech variations are associated with rapid changes of F0, SPL, and DUR patterns, which are influenced by language pragmatics and dialect, but also personal, educational, cultural, social, and occupational aspects. Thus, the adjustments in vocal fold tonus and vocal tract position required by prosodic changes affected the harmonic organization, periodicity, and stability of a voice signal. Our results showed strong effects for CPPS, while the effects were less distinct for HNR. Given this, especially measurements of CPPS in vowels from connected speech are affected by several confounding factors that should be considered in clinical acoustic voice assessments and decision-making. One way to partially control for some confounding factors would be by applying F0 and SPL adapted normative values while assessing SVs and EVs from stressed syllables of speech utterances.

Understanding F0 and SPL effects on CPPS and HNR

In the present study, voice pathologies like polyps, nodules, vocal fold palsy, sulcus, and vocal chink may have increased signal aperiodicity and noise, lowering CPPS/HNR (Akbari, Seifpanahi, Ghorbani, Izadi, & Torabinezhad, 2018; Kumar et al., 2010; Lopes et al., 2017). Also, higher F0 may have been produced with more strained and louder phonation in the present group (Awan et al., 2012; Brockmann-Bauser et al., 2018; Kumar et al., 2010; Teixeira & Fernandes, 2015; Wolfe & Martin, 1997a). Louder and higher voicing involves a higher medial compression of the vocal folds. Such adjustments may result in improved glottal closure and signal periodicity, causing better (higher) CPPS and HNR measures (Awan et al., 2012; Brockmann et al., 2011). Thus, besides prosodic speech variations, general voice use patterns associated with F0 and SPL changes are expected to influence CPPS and HNR.

While the direct relationship of CPPS and HNR with SPL is consistent in healthy and dysphonic voices (Brockmann-Bauser et al., 2018; Phadke et al., 2018), the relationship with F0 seems more variable for dysphonic voices (Lopes et al., 2017; Teixeira & Fernandes, 2015). Also, differences in vocal behavior associated with gender, such as a generally louder speaking voice SPL in men, make the interpretation of the interaction between F0, SPL and acoustic measures more complex (Awan et al., 2012; Awan et al., 2010; Brockmann et al., 2011; Brockmann-Bauser et al., 2018; Hasanvand et al., 2017). In healthy voices, sustained vowels tend to be breathier and less loud in women. However, this is less marked in speech utterances (Awan et al., 2010; Brockmann-Bauser et al., 2018; Hasanvand et al., 2017). Therefore, we recommend that clinicians mind the patient´s vocal behavior during voice recordings in order to obtain more representative samples of the habitual voice performance. This should mainly be applied to obtain more reliable comparisons between assessments before and after therapy.

Interpreting the effects of DUR and syllable stress on CPPS and HNR

In the present study, the shorter sample durations (i.e., vowel length) obtained from extracted vowels resulted in better CPPS and HNR ratings. One possible reason for this may be the smaller content regarding signal periodicity and harmonicity in EVs (65 to 220ms) as compared to SV (500ms).

Concerning CPPS, the short samples may have undergone a partial smoothing process, resulting in higher CPPS scores. Praat program only performs the time-smoothing when two or more frames emerge after the power cepstrogram computation (Boersma & Weenink, 2018). That was the case for 33% of the voice samples, which had a sample duration above 102ms and pitch frames above 23 before smoothing. The other 77% were below 106ms and 24 pitch frames, having only one time-frame after power cepstrogram computation. Based on these results and considering the settings used in our signal processing, we conclude that 110ms and 23 pitch frames would be the minimal requirement for homogeneous and systematic CPP smoothing in Praat. Future studies controlling the smoothing techniques along with window length and F0 should give a clear concept of the clinically most useful methods for smoothing in short signals.

For HNR, the combination of very short windows length (less than 80ms) combined with aperiodic signals and a certain degree of noise may have caused more sample losses and a larger dispersion of values with more outliers. As pointed out by Boersma (1993), HNR values do not deteriorate in very short windows, but there is an increase in pitch determination errors, leading to measurement inaccuracies. Thus, the ideal condition for an accurate short-term HNR measurement in Praat would be at least a minimum of 4.5 fundamental frequency periods and 80ms duration (DUR).

Regarding the effect of syllable stressing patterns, vowels from stressed syllables had a higher mean sample duration and thus were more suitable for short-term analysis of spectral and cepstral-based measures. However, vowels with similar syllable-stressing conditions still may present different F0 and SPL depending on their position in the word and the sentence due to prosodic variations (in press). We observed that stressed EVs 5 (s**A**be) and 7 (l**Á**), both positioned at the beginning of a word and sentence, showed high CPPS, HNR, F0, and SPL means altogether. This observation is compliant with an ascendant prosodic intonation, related to more vocal effort, laryngeal tension, and thereby in our study a higher signal periodicity and harmonicity (Brockmann-Bauser et al., 2018; Park & Stepp, 2018). On the opposite, stressed EV 4 (Acab**A**r), positioned at the end of a word and sentence, showed lower CPPS, HNR, F0, and SPL altogether. Consequently, this may result from lower vocal fold tension, higher glottal open quotients, and more variable mucosa cover vibration in the vocal folds (Brockmann-Bauser et al., 2018).

CPPS versus HNR in clinical diagnostics

In this work, CPPS and HNR showed high correlations with each other in SV and EVs from stressed and pre-stressed syllables. While CPPS showed more stable and robust results, for HNR, there was more loss of samples, outliers, and variability of values. This suggests that, despite some conceptual similarities between CPPS and HNR, CPPS is more precise for measuring aperiodic signals in extracted vowels from speaking samples due to its relative independence of correct pitch estimation (Fraile & Godino-Llorente, 2014; Maryn & Weenink, 2015; Riesgo & Nöth, 2019; Shue et al., 2010). Moreover, HNR is influenced by cycle boundaries tracking and vocal tract configuration (Fraile & Godino-Llorente, 2014; Parsa & Jamieson, 2001; Teixeira & Fernandes, 2015). Adding noise may mask the underlying F0, leading to a higher possibility of pitch determination errors and inaccuracies in HNR rates (Boersma, 1993). In turn, CPPS seemed to be more sensitive to articulatory and speaking prosodic effects, calling attention to the necessity of controlling these factors during voice assessment.

The higher accuracy of CPPS to detect voice disorders using connected speech analysis (Sauder et al., 2017), does not entail a full advantage over HNR in the clinical setting. The clinical application of CPPS to discriminate the degree, type, and etiology of dysphonia is still limited (Phadke et al., 2018; Riesgo & Nöth, 2019). In our study, HNR was less sensitive to changes in articulatory, F0, SPL, and DUR. Further, HNR may be more suitable for acoustic assessment with sustained vowels in type 1 and 2 signals. As argued by several authors, HNR may be equally or more reliable than CPPS to identify dysphonia and roughness in SV (Latoszek, De Bodt, Gerrits, & Maryn, 2018; Latoszek, Maryn, Gerrits, & De Bodt, 2017; Moers et al., 2012; Vaz Freitas et al., 2015). HNR was also incorporated to multiparametric indices such as AVQI and has potentially high applicability in short-time segment analysis in certain types of voice pathology (Klingholtz, 1990; Maryn & Weenink, 2015; Zhang & Jiang, 2008). Additionally, cepstral analysis is not yet widely available in custom software accessible to general health care professionals, unlike HNR. Severely dysphonic voices represent nearly a quarter of all patients assisted in health care facilities (Gillespie, Dastolfo, Magid, & Gartner-Schmidt, 2014), which means that 75% of patients might be suitable for spectral or frequency-based acoustic measurements available in most custom acoustic analysis tools. This should be considered for health care systems with limited resources for clinical assistance.

The research community broadly recognizes the Praat program as a reliable open-source software (Boersma, 2002). Praat has been shown to discriminate perceptual dysphonia using frequency, spectral and cepstral-based measures like HNR and CPPS respectively (Maryn &

Weenink, 2015; Vaz Freitas et al., 2015; Watts et al., 2017), which turns it an attractive tool for the clinical application and research in low-cost technology circumstances. Additionally, Praat allows the concurrent calculation of multiple acoustic measures, enabling the study of complex associations between voice parameters. Moreover, it is highly suitable for short-term analysis, allowing the calculation of both SV and speech samples in different ways. This software uses the Gaussian-shaped window as standard that gives no adjacent lobes and analyzes more samples per frame (Boersma & Weenink, 2018) in comparison to the commonly used rectangular and Hamming window types. However, so far, there is a lack of validated normative values for acoustic measures analyzed with Praat, which hinders the clinical interpretation of measurement results. Given this, future studies should provide normative and standardized measures using Praat program considering the influence of F0, SPL, and token.

Assuming that CPPS and HNR have been mostly compared to perceptual results so far, further studies from the perspective of laryngeal, pulmonary, and articulatory function considering various diagnosis groups would contribute to understanding the advantages of CPPS and HNR in clinical decision making.

Cepstral and spectral measures in sustained and extracted vowels from speech

In our study, we aimed to compare cepstral and spectral values between sustained and extracted vowels from connected speech samples in Brazilian Portuguese language, including a wide variety of ages, voice disorders, and degrees of perceptual dysphonia. We presume that CPPS and HNR results for SV and EV in our study were influenced by the high prevalence of absent or mild perceptual dysphonia (59%) and absent or mild breathiness (92%), along with the exclusion of type 3 signals. However, the relationship between Type 3 and 4 signals and reliability of CPPS measurements have been poorly studied so far.

When comparing our CPPS distribution for EVs (means 22-26.9dB) with other works using Praat, we found similar values on both non-dysphonic (mean range 16.5-23dB) and dysphonic group (14.7-20.3dB) as reported by Sauder et al.(2017), who analyzed voiced and unvoiced components in the second sentence of the rainbow passage. Also, Kitayama et al. (2018) identified thresholds ranging 20.8-23dB for discriminating normophonic from dysphonic voices, using only voiced sentence passages in Japanese language. Moreover, Watts et al. (2017) found a difference of 9 dB in CPPS indices from speaking samples in Finnish and English languages. The proportion of unvoiced components in speaking samples, which varies among languages, may influence the prominence of harmonic organization over the noise content measured by the cepstral parameters, yielding divergent CPPS results among speaking samples (Zhang & Jiang, 2008).

As for sustained vowels, our mean CPPS (23dB) was similar to values reported by Watts et al. (2017) (22.6/22.9dB) in healthy and dysphonic speakers. Further, our results were lower than those found by Kitayama et al. (2018) (29.5/30.3dB) in normophonic speakers, and higher than values reported by Phadke et al. (2018) (13.6 dB), Nuñez-Batalla et al. (2018) (16dB), Delgado-Hernandez et al. (2018) (14.9 dB for normophonic and 11.5dB for dysphonics), and Brockmann-Bauser et al. (in press) (16 dB) in healthy female speakers.

Regarding HNR, our mean value for sustained vowel (18.7dB) was almost like those published by Vaz-Freitas, Pestana, Almeida, & Ferreira (2018) (19.5dB) measured in a heterogeneous group of dysphonia degrees. Further, our results were lower as compared to a group of healthy speakers from studies by Delgado et al. (2017) and Brockmann-Bauser et al. (2018) (respectively 24.9dB and 27.7dB). Further, Moers et al. (2012)found a mean HNR of 12.8dB in a group with varying degrees of dysphonia, and a mean of 15.7 dB in the group with the more reliable pitch recognition. Based on the above-described results from the literature, we conclude that to date, there is no consensus regarding normative values representative of healthy and disordered voices for HNR and CPPS using Praat program. Therefore, in future studies, it is important to control for gender, F0, SPL, type of acoustic signal (token), algorithm, and prevalence of perceptual dysphonia while reporting acoustic measures.

Speech tasks might highlight shortcomings of an individual's voice behavior, such as by an effort and endurance test challenging vocal performance (Klingholtz, 1990). Given this, the analysis of extracted vowels from connected speech provides additional information about vocal fold behavior and compensations in the presence of pragmatic, respiratory, and articulatory demands (Watts & Awan, 2015). Also, comparing acoustic information of stressed vowels in different positions may evidence specific voice disorders. Previous reports have already shown extreme changes in a word section spoken by dysphonic patients, attributed to shifting in phonatory regularity at the termination of a breath group (Watts & Awan, 2015).

The control of smaller short-term speech sections, however, should be applied as complementary to SV, since both contexts may convey clinically relevant and distinct information about vocal behavior in healthy and dysphonic voices (Watts & Awan, 2011). Sustained vowels would emphasize instabilities of F0 and SPL in relatively stable conditions, such as in asymmetry or irregularity of vocal fold vibration, some types of hyperfunctional dysphonia, and essential tremor. Speech tasks could underline signs of severe irregularities of vocal fold vibration or vocal fatigue. Hypofunctional voice disorders, Reinke’s edema, vocal fold scars, and *sulcus vocalis*, for example, may show fewer variations in SPL and F0 during speech utterance, while cerebellar voice disorders may show exacerbated variations or hypernasal resonance.

Conclusions

In dysphonic voices, F0, SPL, and DUR affected more markedly CPPS than HNR. Further, CPPS was affected by vowel context (sustained versus extracted vowel) and syllable stress, while HNR was not influenced by these factors. Thus, CPPS may be more sensitive to changes in vocal tract configuration than HNR. Despite that, CPPS seems more suitable for assessing extracted vowels from speaking samples, while HNR may be adequate in the assessment of dysphonia using sustained vowels, especially when pitch and intensity are correctly recognized. Further studies of cepstral and spectral measures in pathologic voices are warranted to understand how these influencing factors may be controlled for in measurements of CPPS and HNR. Controlling the smoothing techniques along with window length and F0 should give a clear concept of the clinically most useful methods on a short-term basis.

Acknowledgements

The authors have no conflicts of interest to disclosure. We want to thank the speech therapist Yara Pirajá Faria for calibrating the microphone and the Praat intensity-SPL calibration, and thanks to the undergraduate students Maria Rita Marques Almeida Teles and Marcelo Santos de Souza for helping with the retrospective data collection. Thanks to the students of the Scientific Initiation Program from UFBA Alicia da Conceição Silva, Ruan Carlos Pereira Borges Nascimento, Felipe Franklin Souza Santa Rosa Mascarenhas for providing support on the auditory-perceptual analysis.

References

Aichinger, P., & Kubin, G. (2018). Re: Gaskill CS, Awan JA, Watts CR, Awan SN. Acoustic and perceptual classification of within-sample normal, intermittently dysphonic, and consistently dysphonic voice types. J Voice. 2016;31:218-228. *Journal of Voice, 32*(3), 381-382. doi:10.1016/j.jvoice.2017.05.023

Akbari, E., Seifpanahi, S., Ghorbani, A., Izadi, F., & Torabinezhad, F. (2018). The effects of size and type of vocal fold polyp on some acoustic voice parameters. *Iranian Journal of Medical Sciences, 43*(2), 158-163.

Audacity, T. (2018). Audacity® software is copyright © 1999-2019 (Version 2.2.2) [Acessed in 15th November 2018]: Free software distributed under the terms of the GNU General Public License. The name Audacity® is a registered trademark of Dominic Mazzoni. https://audacityteam.org/. Retrieved from https://audacityteam.org/

Awan, S. N., Awan, J. A., Watts, C. R., & Gaskill, C. S. (2018). Response to Aichinger and Kubin Re: Letter to the Editor "Acoustic and Perceptual Classification of Within-Sample Normal, Intermittently Dysphonic, and Consistently Dysphonic Voice Types". *Journal of Voice, 32*(3), 383-384. doi:10.1016/j.jvoice.2017.06.001

Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of Vocal Intensity and Vowel Type on Cepstral Analysis of Voice. *Journal of Voice, 26*(5). doi:ARTN 670.e1510.1016/j.jvoice.2011.12.001

Awan, S. N., Roy, N., Jette, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24*(9), 742-758. doi:10.3109/02699206.2010.492446

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice* (2nd ed.): Delmar Learning.

Barbosa, P. A., & Madureira, S. (2015). *Manual de fonética acústica experimental: aplicações a dados do português. [Manual of experimental acoustic phonetics: applications to Portuguese data]*. São Paulo, Brasil: Cortez Editora.

Behlau, M. (2004). Consensus auditory-perceptual evaluation of voice (CAPE-V), ASHA 2003. *Revista da Sociedade Brasileira de Fonoaudiologia, 9*(3), 187-189.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the institute of phonetic sciences, 17*(1193), 97-110.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international, 5*(9/10), 341-345.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. (Version 6.0.46) [Acessed in 15th November 2018]: http://www.praat.org/. Retrieved from http://www.praat.org/

Brinca, L. F., Batista, A. P., Tavares, A. I., Goncalves, I. C., & Moreno, M. L. (2014). Use of cepstral analyses for differentiating normal from dysphonic voices: a comparative study of connected speech versus sustained vowel in European Portuguese female speakers. *Journal of Voice, 28*(3), 282-286. doi:10.1016/j.jvoice.2013.10.001

Brockmann, M., Drinnan, M. J., Storck, C., & Carding, P. N. (2011). Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *Journal of Voice, 25*(1), 44-53. doi:10.1016/j.jvoice.2009.07.002

Brockmann-Bauser, M., Bohlender, J. E., & Mehta, D. D. (2018). Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders. *Journal of Voice, 32*(2), 162-168. doi:10.1016/j.jvoice.2017.04.008

Brockmann-Bauser, M., Van Stan, J. H., Sampaio, M. C., Bohlender, J. E., Hillman, R. E., & Mehta, D. D. (in press). Effects of vocal intensity and fundamental frequency on cepstral peak prominence in patients with voice disorders and vocally healthy controls. *Journal of Voice*.

Castellana, A., Carullo, A., Astolfi, A., Bisetti, M. S., & Colombini, J. (2018). Vocal health assessment by means of Cepstral Peak Prominence Smoothed distribution in continuous speech. *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 7. doi:10.1109/MeMeA.2018.8438655

Castellana, A., Selamtzis, A., Salvi, G., Carullo, A., & Astolfi, A. (2017). Cepstral and entropy analyses in vowels excerpted from continuous speech of dysphonic and control speakers. *18th Annual Conference of the International Speech Communication Association (Interspeech 2017), 1-6*, 1814-1818. doi:10.21437/Interspeech.2017-335

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*: Pearson Education.

Delgado, J., León, N. M., Jiménez, A., & Izquierdo, L. M. (2017). Análisis acústico de la voz: medidas temporales, espectrales y cepstrales en la voz normal con el Praat en una muestra de hablantes de español. *Revista de Investigación en Logopedia, 7*(2), 108-127.

Delgado-Hernandez, J., Leon-Gomez, N. M., Izquierdo-Arteaga, L. M., & Llanos-Fumero, Y. (2018). Cepstral analysis of normal and pathological voice in Spanish adults. Smoothed cepstral peak prominence in sustained vowels versus connected speech. *Acta Otorrinolaringologica Española, 69*(3), 134-140. doi:10.1016/j.otorri.2017.05.006

Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice, 19*(1), 15-28. doi:https://doi.org/10.1016/j.jvoice.2004.07.003

do Amaral Catani, G. S., Hamerschmidt, R., Moreira, A. T., Timi, J. R. R., Wiemes, G. R. M., Ido, J., & Macedo, E. (2016). Subjective and objective analyses of voice improvement after phonosurgery in professional voice users. *Medical Problems of Performing Artists, 31*(1), 18-24. doi:http://dx.doi.org/10.21091/mppa.2016.1004

Fraile, R., & Godino-Llorente, J. I. (2014). Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control, 14*, 42-54. doi:https://doi.org/10.1016/j.bspc.2014.07.001

Gerratt, B. R., Kreiman, J., & Garellek, M. (2016). Comparing measures of voice quality from sustained phonation and continuous speech. *Journal of Speech Language and Hearing Research, 59*(5), 994-1001. doi:10.1044/2016_Jslhr-S-15-0307

Gillespie, A. I., Dastolfo, C., Magid, N., & Gartner-Schmidt, J. (2014). Acoustic analysis of four common voice diagnoses: moving toward disorder-specific assessment. *Journal of Voice, 28*(5), 582-588. doi:10.1016/j.jvoice.2014.02.002

Halberstam, B. (2004). Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL Journal for Oto-rhino-laryngology and its Related Specialties, 66*(2), 70-73. doi:10.1159/000077798

Hasanvand, A., Salehi, A., & Ebrahimipour, M. (2017). A cepstral analysis of normal and pathologic voice qualities in Iranian adults: A comparative study. *Journal of Voice, 31*(4). doi:UNSP 508.e17 10.1016/j.jvoice.2016.10.017

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., . . . Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology Rhinology and Laryngology, 112*(4), 324-333. doi:10.1177/000348940311200406

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*(1), 20-27.

Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., . . . Hillenbrand, J. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice, 28*(6). doi:10.1016/j.jvoice.2014.05.005

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech Language and Hearing Research, 39*(2), 311-321.

Hirano, M. (1981). GRBAS scale for evaluating the hoarse voice & frequency range of phonation. *Clinical examination of voice, 5*, 83-84.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45*, 326-347.

Kitayama, I., Hosokawa, K., Iwahashi, T., Iwahashi, M., Iwaki, S., Kato, C., . . . Inohara, H. (2018). Intertext variability of smoothed cepstral peak prominence, methods to control it, and its diagnostic properties. *Journal of Voice*. doi:10.1016/j.jvoice.2018.09.021

Klingholtz, F. (1990). Acoustic recognition of voice disorders - A comparative-study of running speech versus sustained vowels. *Journal of the Acoustical Society of America, 87*(5), 2218-2224. doi:Doi 10.1121/1.399189

Koufman, J. A., & Isaacson, G. (1991). The spectrum of vocal dysfunction. *Otolaryngologic Clinics of North America, 24*(5), 985-988.

Kumar, B. R., Bhat, J. S., & Prasad, N. (2010). Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice, 24*(6), 651-653. doi:10.1016/j.jvoice.2009.07.008

Latoszek, B. B. V., De Bodt, M., Gerrits, E., & Maryn, Y. (2018). The exploration of an objective model for roughness with several acoustic markers. *Journal of Voice, 32*(2), 149-161. doi:10.1016/j.jvoice.2017.04.017

Latoszek, B. B. V., Maryn, Y., Gerrits, E., & De Bodt, M. (2017). The acoustic breathiness index (ABI): A multivariate acoustic model for breathiness. *Journal of Voice, 31*(4), 511-e511. doi:https://doi.org/10.1016/j.jvoice.2016.11.017

Lopes, L. W., Simoes, L. B., Silva, J. D., Evangelista, D. S., da Nobrega, E., C., U. A., . . . Vieira, V. J. D. (2017). Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses. *Journal of Voice, 31*(3), 382 e315-382 e326. doi:10.1016/j.jvoice.2016.08.015

Lowell, S. Y., & Hylkema, J. A. (2016). The effect of speaking context on spectral- and cepstral-based acoustic features of normal voice. *Clinical Linguistics & Phonetics, 30*(1), 1-11. doi:10.3109/02699206.2015.1087049

Madill, C., Nguyen, D. D., Cham, K. Y., Novakovic, D., & McCabe, P. (2019). The impact of nasalance on cepstral Peak prominence and h armonics-to-noise ratio. *The Laryngoscope, 129*, E299–E304. doi:10.1002/lary.27685

Maryn, Y., & Weenink, D. (2015). Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice, 29*(1), 35-43. doi:10.1016/j.jvoice.2014.06.015

Moers, C., Mobius, B., Rosanowski, F., Noth, E., Eysholdt, U., & Haderlein, T. (2012). Vowel- and text-based cepstral analysis of chronic hoarseness. *Journal of Voice, 26*(4), 416-424. doi:10.1016/j.jvoice.2011.05.001

Nuñez-Batalla, F., Cartón-Corona, N., Vasile, G., García-Cabo, P., Fernández-Vañes, L., & Llorente-Pendás, J. L. (2018). Validez de las medidas del pico cepstral para la valoración objetiva de la disfonía en sujetos de habla hispana. *Acta Otorrinolaringológica Española, In Press*, 7. doi:https://doi.org/10.1016/j.otorri.2018.04.008

Park, Y., & Stepp, C. E. (2018). The Effects of stress type, vowel identity, baseline F0, and loudness on the relative fundamental frequency of individuals with healthy voices. *Journal of Voice, In Press*. doi:https://doi.org/10.1016/j.jvoice.2018.04.004

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. *Journal of Speech Language and Hearing Research, 44*(2), 327-339. doi:10.1044/1092-4388(2001/027)

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., . . . Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology, 27*(3), 887-905. doi:10.1044/2018_ajslp-17-0009

Phadke, K. V., Laukkanen, A. M., Ilomaki, I., Kankare, E., Geneid, A., & Svec, J. G. (2018). Cepstral and Perceptual Investigations in Female Teachers With Functionally Healthy Voice. *Journal of Voice*. doi:10.1016/j.jvoice.2018.09.010

Riesgo, C. A. F., & Nöth, E. (2019). What Makes the Cepstral Peak Prominence Different to Other Acoustic Correlates of Vocal Quality? *Journal of Voice*. doi:https://doi.org/10.1016/j.jvoice.2019.01.004

Rosen, C. A., & Murry, T. (2000). Nomenclature of voice disorders and vocal pathology. *Otolaryngologic Clinics of North America, 33*(5), 1035-+. doi:10.1016/S0030-6665(05)70262-0

Sampaio, M. C., Bohlender, J. E., & Brockmann-Bauser, M. (in press). Fundamental frequency and intensity effects on cepstral measures in vowels from connected speech of speakers with voice disorders. *Journal of Voice*.

Sauder, C., Bretl, M., & Eadie, T. (2017). Predicting Voice Disorder Status From Smoothed Measures of Cepstral Peak Prominence Using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice, 31*(5), 557-566. doi:10.1016/j.jvoice.2017.01.006

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. 607-610.

Selamtzis, A., Castellana, A., Salvi, G., Carullo, A., & Astolfi, A. (2019). Effect of vowel context in cepstral and entropy analysis of pathological voices. *Biomedical Signal Processing and Control, 47*, 350-357. doi:10.1016/j.bspc.2018.08.021

Shevchenko, T. I. (2003). Prosody as a community code (Russian, British and American women in radio talk). *Proceedings of the 15th International Congress of Phonetic Sciences, 3*(9), 1189-1192.

Shue, Y. L., Chen, G., & Alwan, A. (2010). On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. *Eleventh Annual Conference of the International Speech Communication Association.*, 34-37.

Skowronski, M. D., Shrivastav, R., & Hunter, E. J. (2015). Cepstral peak sensitivity: A theoretic analysis and comparison of several implementations. *Journal of Voice, 29*(6), 670-681. doi:10.1016/j.jvoice.2014.11.005

Teixeira, J. P., & Fernandes, P. O. (2015). Acoustic analysis of vocal dysphonia. *Procedia Computer Science, 64*, 466-473. doi:10.1016/j.procs.2015.08.544

Titze, I. R. (1995). *Workshop on Acoustic Voice Analysis: Summary Statement*, National Center for Voice and Speech, Iowa.

Vaz Freitas, S., Melo Pestana, P., Almeida, V., & Ferreira, A. (2015). Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *J Voice, 29*(3), 390 e391-397. doi:10.1016/j.jvoice.2014.08.007

Vaz-Freitas, S., Pestana, P. M., Almeida, V., & Ferreira, A. (2018). Acoustic analysis of voice signal: Comparison of four applications software. *Biomedical Signal Processing and Control, 40*, 318-323. doi:https://doi.org/10.1016/j.bspc.2017.09.031

Vogel, A. P., & Maruff, P. (2008). Comparison of voice acquisition methodologies in speech research. *Behavior Research Methods, 40*(4), 982-987. doi:10.3758/brm.40.4.982

Wagener, K., Kühnel, V., & Kollmeier, B. (1999). Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. *Zeitschrift Fur Audiologie, 38*, 4-15.

Watts, C. R., & Awan, S. N. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech Language and Hearing Research, 54*(6), 1525-1537. doi:10.1044/1092-4388(2011/10-0209)

Watts, C. R., & Awan, S. N. (2015). An examination of variations in the cepstral spectral index of dysphonia across a single breath group in connected speech. *Journal of Voice, 29*(1). doi:10.1016/j.jvoice.2014.04.012

Watts, C. R., Awan, S. N., & Maryn, Y. (2017). A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice, 31*(3), 387 e381-387 e310. doi:10.1016/j.jvoice.2016.09.012

Winholtz, W. S., & Titze, I. R. (1997). Conversion of a head-mounted microphone signal into calibrated SPL units. *Journal of Voice, 11*(4), 417-421.

Wolfe, V., & Martin, D. (1997a). Acoustic correlates of dysphonia: type and severity. *Journal of Communication Disorders, 30*(5), 403-415; quiz 415-406.

Wolfe, V., & Martin, D. (1997b). Acoustic correlates of dysphonia: type and severity. *J Commun Disord, 30*(5), 403-415.

Zhang, Y., & Jiang, J. J. (2008). Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *Journal of Voice, 22*(1), 1-9. doi:10.1016/j.jvoice.2006.08.003

Tables and Figures

Table 1. The Brazilian Portuguese version of four CAPE-V sentences: Target vowels /a/ (underlined, in bold and capital letters) and phonetic dialect transcription

Sentence	Phonetic transcription	Extracted vowels
"'**A**gora é hor**A** de ac**A**b**A**r"	a.gˈɔ.ɾ ˈɛ ˈɔ.ɾɐ dʒi a.ka.bˈa	EV1-4
"sônia s**A**be samb**A**r sozinha"	sˈõ.ɲɐ sˈa.bi sã.bˈa sɔ.zˈi.ɐ	EV5, EV6
" olha l**Á** o **A**vião **A**zul"	ˈɔ.λɐ lˈa u a.vi.ˈãu a.zˈu	EV7-9
"p**A**p**A**i trouxe pipoca quente"	pa. pˈai tr ˈo.ʃi pi.pˈɔ.kɐ kˈẽ.tʃi	EV10, EV11

Table 2. Description of acoustic measures and commands used in Praat

Measure	Unit	Description [a, b]	Commands in Praat [b, c]
DUR	s	Reflects the window length of the extracted vowels	1.Sound object: Query time domain; 2.Get total duration.
F0	Hz	Main waveform repetition rate (cycles) per second and forward cross-correlation method (experimental for short-time windows)	1.Sound object: Analyze Periodicity; 2.To Pitch (cc) (standard settings); 3.Pitch object: Get mean.
SPL	dB(A)	Represents the power of an acoustic signal and affects the amplitude of cycles	1.Sound object: To Intensity (standard settings); 2.Intensity object: Get mean, dB method.
HNR	dB	The ratio between harmonic and noise components based on cross-correlation method	1.Sound object: Analyze Periodicity; 2.To Harmonicity (cc) (standard settings); 3.Harmonicity object: Get mean.
CPPS	dB	CPP: Difference in amplitude between the cepstral peak and the corresponding value on the regression line that is directly below the peak CPPS: the Power Cepstrogram object is smoothed by averaging cepstra across time first, and then across quefrency	1.Sound object: Filter (stop Hann Band). Settings: 0 to 34Hz, Smoothing 0.1 Hz; 2.Filtered sound object: To Power Cepstrogram (standard settings); 3.Power Cepstrogram object: Get CPPS. Settings: No subtracting tilt, time window 0.01(5 frames), quefrency window 0.001(10 bin), Peak search range 60-330 Hz, Tolerance 0.05, Parabolic Interpolation, Tilt line frequency range 0.001-0, Straight line, Robust fit method.

dB = decibels, Hz = Hertz, s = seconds. [a] (Baken & Orlikoff, 2000), [b] (Boersma & Weenink, 2018),[c] (Maryn & Weenink, 2015)

Table 3. Distribution of acoustic measures by token type

	SV	EV1	EV2	EV3	EV4	EV5	EV6	EV7	EV8	EV9	EV10	EV11
N =	27	23[a]/22[b]	16	15	15	24	23[a]/22[b]	23[a]/22[b]	14[a]/13[b]	15	15[a]/14[b]	19
Syllable Stress	-	Pre-stressed	Post-stressed	Pre-stressed	Stressed	Stressed	Stressed	Stressed	Pre-stressed	Pre-stressed	Pre-stressed	Stressed
CPPS (dB)		*	*	*	*	*	*	*	*	*	*	*
Mean (SD)	23.1 (2)	24.5 (5)	26.2 (3)	25.9 (2)	22.5 (4)	23.8 (3)	23.2 (4)	23.4 (4)	26.9 (4)	25.9 (4)	23.7 (4)	24.8 (4)
CI	22 - 24	22-27	25-28	25-27	20-25	22-25	22-25	22-25	25-29	24-28	21-26	23-27
HNR (dB)												
Mean (SD)	18.7 (4)	13.4 (4)	17.7 (4)	15.5 (5)	13.4 (6)	15.7 (4)	16.2 (4)	17.1 (5)	20.2 (4)	18.4 (5)	13.6 (6)	15.0 (6)
CI	17 - 20	12 - 15	17-20	13-18	10-16	14-18	15-18	15-19	18-23	16-21	10-17	12-18
F0 (Hz)												
Mean (SD)	165.1(36)	155.1(37)	168.1(32)	172.2(42)	136.6(31)	159.4(37)	148.2(34)	146.9(37)	178.3(30)	160 (42)	159.8(38)	170 (33)
CI	151-180	139-171	151-185	149-195	120-154	144-175	133-163	131-163	160-196	137-183	137-183	154-185
SPL (dBA)												
Mean (SD)	87.2 (3)	88.6 (3)	89.0 (3)	89.6 (3)	86.5 (3)	90.2 (3)	88.8 (3)	89.2 (4)	90.6 (3)	88.0 (3)	81.0 (3)	90.5 (2)
CI	86 - 89	87-90	87-91	88-91	85-88	87-92	88-90	88-91	89-92	86-90	79-83	89-92
DUR (s)												
Mean (SD)	0.50 (.00)	0.08 (.02)	0.07 (.00)	0.08 (.02)	0.10 (.03)	0.12 (.04)	0.09 (.03)	0.12 (.04)	0.07 (.01)	0.07 (.01)	0.07 (.01)	0.08 (.02)
CI	0.50-0.50	0.08-0.09	0.07-0.07	0.06-0.09	0.08-0.11	0.10-0.14	0.08-0.11	0.10-0.14	0.07-0.07	0.07-0.08	0.07-0.08	0.07-0.09

SD =standard deviation, CI = confidence interval, dB = decibels, Hz = Hertz, s = seconds. * ANCOVA with Bonferroni correction shows significant difference between SV and EVs, $p \leq .05$, a = values for CPPS, b = values for HNR.

Table 4. Paired correlation between SV and EV for CPPS, HNR, F0 and SPL

SV versus:	CPPS (dB)	HNR (dB)	F0 (Hz)	SPL (dBA)	*N*
EV 1	0.49*	0.31	0.71**	0.14	23
EV 2	0.45	0.53*	0.49*	0.48	16
EV 3	0.32	0.14	0.43	0.30	15
EV 4	0.41	0.34	0.50	0.17	15
EV 5	0.57*	0.46*	0.72**	0.20	24
EV 6	0.81**	0.42*	0.45*	0.45*	23
EV 7	0.56*	0.20	0.00	0.28	23
EV 8	0.71*	0.38	0.35	0.52	14
EV 9	0.24	0.22	0.25	-0.20	15
EV 10	0.59*	0.55*	0.48	-0.02	15
EV 11	0.45*	0.44	0.64*	-0.09	19

dB = decibels, Hz = Hertz, * Paired sample Pearson´s test shows significant correlation with SV $p \leq .05$, $**p \leq 001$.

Table 5. Effects of F0, SPL, DUR, syllable stress and vowel context on CPPS and HNR

Influencing Factors	CPPS			HNR		
	F	SER	df	F	SER	df
F0 (Hz)	37.9**	0.01	209	27.1**	0.01	204
SPL (dBA)	31.2**	0.07	212	15.8**	0.09	209
DUR (s)	8.9*	6.24	195	5.4*	8.64	187
Syllable Stress	6.9*	2.29	201	1.7	3.80	195
Extracted vowel	2.7*	2.69	196	5.7	3.73	187

CI = confidence interval, SER = standard error, df = degree of freedom, dB = decibels, Hz = Hertz, s = seconds. *ANCOVA significant effect $p \leq .05$, ** ANCOVA highly significant effect $p \leq 001$.

Table 6. Estimated Means of CPPS and HNR after excluding the influence of confounding factors F0, SPL and DUR[a]

Vowel Context	CPPS	HNR
	Mean (CI)	Mean (CI)
SV	30 (26-35)	12 (5-18)
EV 1	23.6 (22-25)	14.1 (14-17)
EV 2	24.1 (22-26)	18.3 (16-21)
EV 3	23.5 (22-25)	15.6 (13-18)
EV 4	23.3 (22-25)	15.8 (14-18)
EV 5	22.9 (22-24) *	15.5 (14-17)
EV 6	23.1 (22-24)	17.2 (15-19)
EV 7	23.2 (22-25)	17.2 (16-19)
EV 8	24.3 (23-26)	19.3 (17-22)
EV 9	24.4 (23-26)	19.2 (18-22)
EV 10	22.3 (21-24)	14.3 (12-17)
EV 11	22.5 (21-24)	15.2 (13-17)

CI = confidence interval, * Bonferroni adjustment for multiple comparisons - significant difference from SV $p \leq .05$, a = Covariates appearing in the model are evaluated at the following values: Mean SPL = 89 dB(A), Mean F0 = 159 Hz, DUR = 0.14s.

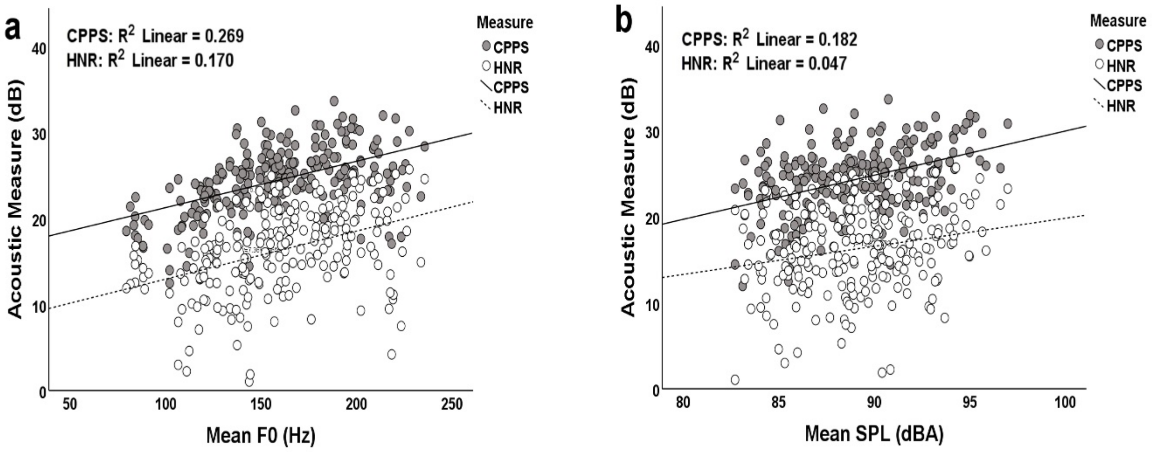

Figure 1. Distribution of CPPS and HNR with F0 (a) and SPL (b). There was a positive relationship for F0 and SPL with CPPS and HNR respectively.

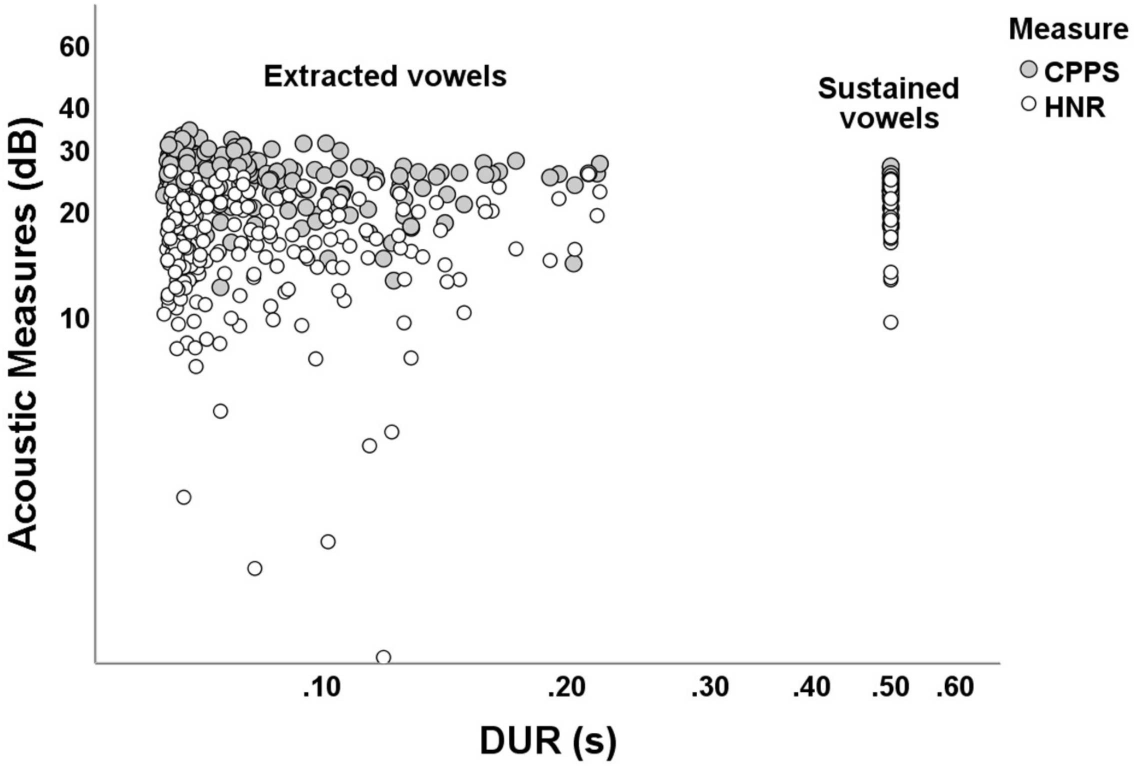

Figure 2. Distribution of CPPS and HNR with DUR. Logarithmic transformation was used to report the distribution of the acoustic measure (*Y*-axis) and sample duration (*X*-axis). In general, shorter samples showed better CPPS values.

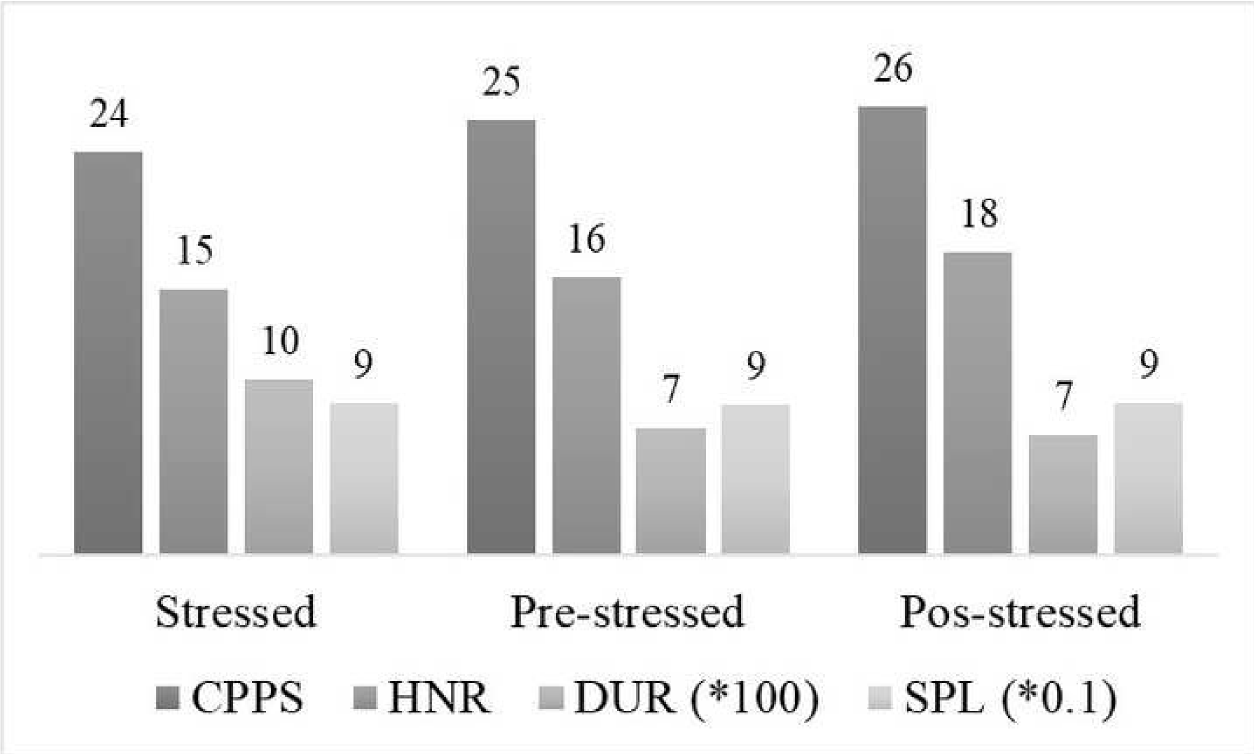

Figure 3. Distribution of CPPS (dB), HNR (dB), DUR (seconds, values were multiplied by 100), and SPL (dB, values were multiplied by 0.1) across type of syllable stress. The higher the vowel duration, the lower the acoustic values (CPPS and HNR)